

# Open Research Online

---

The Open University's repository of research publications  
and other research outputs

## Data challenges for the Gaia Science Alerts System

### Conference or Workshop Item

#### How to cite:

Burgon, Ross; Wyrzykowski, Lukasz and Hodgkin, Simon (2010). Data challenges for the Gaia Science Alerts System. In: Astronomical Data Analysis Software and Systems XX ASP Conference Series, 7-11 Nov 2010.

For guidance on citations see [FAQs](#).

© 2010 The Authors

Version: Accepted Manuscript

Link(s) to article on publisher's website:  
<http://adass2010.cfa.harvard.edu/>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

## Data Challenges for the Gaia Science Alerts System

Ross Burgon<sup>1</sup>, Lukasz Wyrzykowski<sup>2</sup> and Simon Hodgkin<sup>2</sup>

<sup>1</sup>*Planetary and Space Sciences Research Institute, The Open University,  
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom*

<sup>2</sup>*Institute of Astronomy, University of Cambridge, Madingley Road,  
Cambridge, CB3 0HA, United Kingdom*

**Abstract.** Gaia is a European Space Agency (ESA) cornerstone mission due to launch late 2012. Its mission is to precisely survey over one billion sources to create an accurate three-dimensional map of the sky. The Gaia Science Alerts (GSA) System, based in the Institute of Astronomy (IoA) at Cambridge University in the UK, aims to use the daily data stream from Gaia to look for and report on transient events both from within and beyond our galaxy. The data stream will be processed in near real-time in order to provide rapid alerts to facilitate ground-based follow-up. This paper provides an overview of the Gaia Science Alerts System and highlights the data processing and storage challenges from data ingestion and event-detection to event classification and the eventual publication mechanism.

### 1. Introduction

ESA's Gaia mission aims to create a 3-dimensional map of the sky to a parallax accuracy  $<7\mu\text{as}$ , a photometric accuracy  $<10\text{mmag}$  and a radial velocity accuracy of  $<1\text{km s}^{-1}$  (Perryman et al. 2001). In order to achieve this Gaia will utilize the largest CCD focal plane ever to be flown in space (Gare et al. 2010). Over 100 individual CCDs will make up the astrometric, photometric and radial velocity instruments on the spacecraft resulting in over a billion pixels that, when read in time-delay integration (TDI) mode will create a data rate of 7 Mb per second.

Gaia will perform its observations from the L2 Lagrange point of the Sun-Earth system. It will operate for 5 years, spinning constantly at 60 arcseconds/sec. The two astrometric fields of view will scan across all objects located along a great circle perpendicular to the spin axis. Both fields of view will be registered on one focal plane, but (almost) the same part of the sky will transit the second FOV 106.5 minutes after the first. The spin axis precesses slowly on the sky resulting in multiple observations of the whole sky over the lifetime of the spacecraft. For a spin rate of 60 arcseconds/sec and a solar aspect angle of 45 degrees, the precessional period will equal 63 days.

The data rate transmitted to the ground stations will be approximately 50 GB per day. Pipeline processing for this data will be managed by the Gaia Data Processing and Analysis Consortium (DPAC), a European-wide organization tasked with creating the final Gaia catalog by 2021.

## 2. Gaia Science Alerts

The Gaia Science Alerts (GSA) System is based at the Institute of Astronomy (IoA) at the University of Cambridge in the UK. The aims of the system are to:

1. detect unexpected and rapid changes in the flux, spectrum or the position or appearance of new objects
2. release alerts to the community and trigger ground-based follow-up
3. monitor a ‘watch list’ of known potentially eruptive sources

The GSA system will achieve this by running in near real time, using the photometric, spectroscopic and astrometric data available from the Gaia data stream and cross-matching sources against existing information. As Gaia is primarily an astrometry mission the design presents many obstacles to performing transient detection and classification. The GSA system will therefore be performed on a best effort basis and as such has only limited manpower and hardware resources.

### 2.1. Triggers, Contaminants and Cadence

Objects of interest to the GSA system include supernovae, micro-lensing events, GRB afterglows, M-dwarf flares, R CrB-type stars, classical and recurrent novae and unknown transients. However there will be triggers, including variable stars, AGNs and asteroids that will be viewed as contaminants as other units within DPAC are tasked with processing these objects (O’Mullane et al. 2007). Figure 1 illustrates some of the types of objects of interest defined in the magnitude change / transient duration parameter-space accessible by Gaia. The large range in transient duration detection is partly due to the design of the focal plane and partly due to the Gaia scanning law (Jordan 2008).

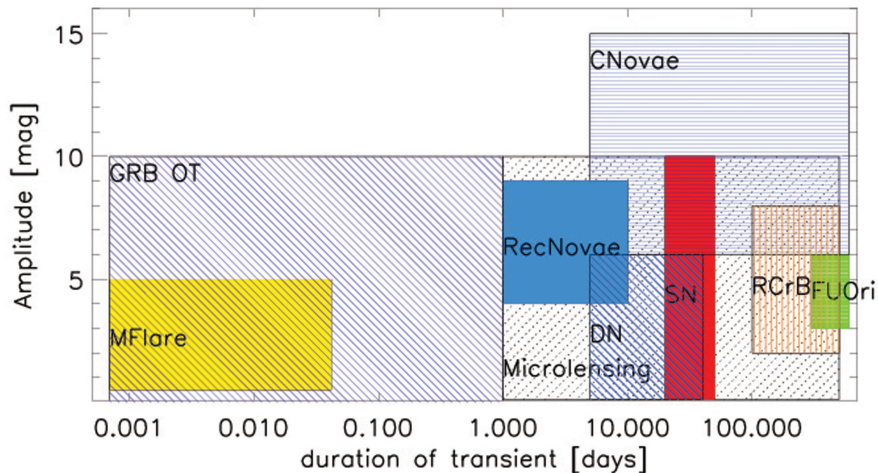


Figure 1. Illustration of the objects that reside within the magnitude change / transient duration parameter space potentially accessible by Gaia

## 2.2. Science Alerts Pipeline

The processing pipeline (AlertPipe) is illustrated in Figure 2. New data will arrive at the pipeline having already been preprocessed at the Science Operations Centre in Madrid with an expected mean rate of 13.5 million observations per day. The data may only be a few hours old at this point but could be up to 36 hrs old. It is then fed into the pipeline where it is ingested and processed.

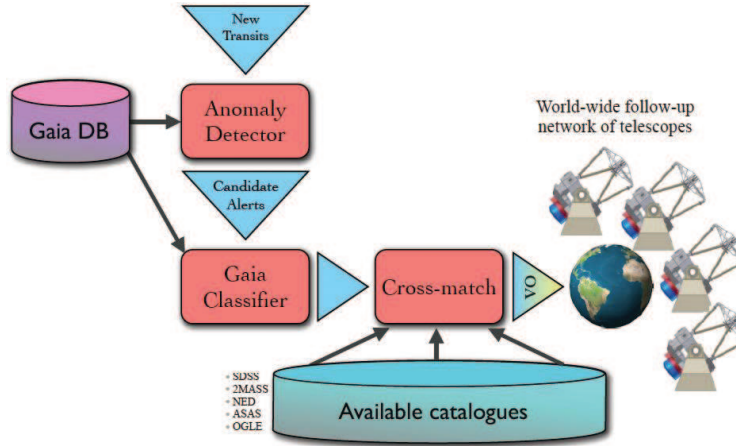


Figure 2. Illustration of the Gaia Science Alerts processing pipeline.

### 2.2.1. Data Storage

The end of mission data volume to be assessable by AlertPipe could be up to 100 TB so it is essential that the storage model is robust and efficient enough to operate within a real-time processing environment. The current implementation for the data storage involves multiple PSQL databases. The entire sky is split into HealPix(els) (Gòrski et al. 2005) and a separate database is used for all sources within each HealPix. This sub-division of the data storage model is to control the size of each the database (which will be subjected to a large number of I/O operations) and should aid multi-threading within AlertPipe (though only a single-threaded implementation is used at present).

### 2.2.2. Anomaly Detector

The detection will be two-fold: (i) known sources exhibiting anomalous behavior and (ii) the appearance of new sources. Both types of alert detection will rely solely on the fluxes observed. Anomaly detection with an ‘old source’ will rely on the history of that source, ideally incorporating both Gaia observations collected so far and historic ground-based data. New sources are defined as not existing in the Gaia database despite previous observation. Naturally, triggers on new sources will only be valid after enough data is gathered by the mission and the whole sky observed at least once or twice.

### 2.2.3. Object Classification and Cross-matching

Classification of objects that are flagged by the anomaly detector will be achieved in three ways: (i) light curve classification, (ii) spectral classification and (iii) position cross-matching with existing catalog sources.

- Classification on the morphology of the light curve will be performed using historical Gaia observations of the source and the new observations that tripped the anomaly detector. The current implementation for this algorithm is based on the Bayesian Classifier with Gaussian Mixture technique Debosscher et al. (2007).
- Spectral classification will be based on the spectrum at the epoch the anomaly detector was tripped. The current implementation for this is through a Self Organized Map (SOM).
- Cross match classification involves data-mining astronomical catalogs around the observed position of the source and identifying any links between the Gaia observation and catalog observations. Catalog access is currently through ‘http get’ requests however it is likely that local copies of each catalog will be maintained at the IoA to decrease network latency issues.

#### 2.2.4. Alert Publication

The alerts will be published in the VOEvent format which offers a standardized but flexible way to convey alert information that is both human and machine-readable. Alert dissemination will be based on three mechanisms. The primary mechanism will be through a subscription e-mail service offered by the IoA which simply e-mails the VOEvent to subscribers. Alert information will also be available via the Gaia Science Alerts website, that in addition to the VOEvent, will also offer ancillary information and images/charts for each alert. Finally, it is intended to stream VOEvent to a VOEvent repository where it can be used for data-mining through services like SkyAlert.org.

### 3. Further Information

Interested readers are directed to the wiki<sup>1</sup> for further information and contact details.

#### References

- Debosscher, L., Sarro, L., Aerts, C., Cuypers, J., Vandenbussche, B., & Solano, E. 2007, *Astronomy and Astrophysics*, 475, 1159
- Gare, P., Sarri, G., & Schmidt, R. 2010, *ESA Bulletin*, 137, 51
- Górski, K., Hivon, E., Banday, A., Wandelt, B., Hansen, F., Reinecke, M., & Bartelmann, M. 2005, *The Astrophysical Journal*, 622, 759
- Jordan, S. 2008, *Astronomische Nachrichten*, 329, 875
- O’Mullane, W., Lammers, U., Bailer-Jones, C., Bastian, U., Brown, A., Drimmel, R., Eyer, L., Huc, C., Katz, D., Lindegren, L., Pourbaix, D., Luri, X., Torra, J., Mignard, F., & van Leeuwen, F. 2007, vol. 376 of *Astronomical Data Analysis Software and Systems XVI ASP Conference Series*, 99
- Perryman, M., de Boer, K., Gilmore, G., Hog, E., Lattanzi, M., Lindegren, L., Mignard, F., Pace, O., & de Zeeuw, P. 2001, *Astronomy and Astrophysics*, 369, 339

---

<sup>1</sup><http://www.ast.cam.ac.uk/research/gsawg/>